Introducing Second Language Assessment
Gary J. Ockey. Cambridge: Cambridge University Press. 2025. Pp. xxiii + 240.

Published by Cambridge University Press in 2025, *Introducing Second Language Assessment* by Gary J. Ockey, Professor of Applied Linguistics at Iowa State University, is a recent addition to the field of second language (L2) assessment. The book is aimed at educators, researchers, and students who are interested in gaining a general understanding of how language assessment works in theory and practice.

Consisting of 13 chapters, the book is divided into five parts, each dealing with a core topic in language assessment. The first part titled "Engaging with Language Assessments" contains two chapters that introduce the foundations of L2 assessment. Chapter 1 differentiates between testing and assessment, elucidates how these may be misused in different contexts, and describes their effective uses. Chapter 2 presents various types of tests and assessments with relevant examples. It encourages readers to reflect on their own experiences with language testing and illustrates how assessments can influence educational outcomes, immigration policies, and employment opportunities. This reflective tone sets a learner-centered perspective that continues for the remainder the book.

The second part titled "Gaining Knowledge of Language Assessment Principles" consists of three chapters. In Chapter 3, Ockey discusses various contextual factors that influence assessment procedures, including sociopolitical, cultural and educational ones, and how these influence stakeholders and language learning policies. Chapters 3 and 4 unpack the assessment principles of validity, uniformity, and reliability. Validity is explained in relation to linguistic theory, needs analysis, corpora and curriculum objectives, while uniformity is explored through practices like content and scoring procedures that can lead to more reliable testing outcomes. The section concludes with an overview of basic approaches to reliability, including test-retest, parallel forms, and internal reliability. Throughout these chapters, the author integrates examples and references to real-world assessments to ground these principles in practice.

The third part of the book "Right or Wrong" consists of four chapters that address issues pertaining to scoring. Chapter 6 focuses on dichotomously scored items and their use in assessing listening and reading comprehension, as well as language subskills such as grammar and vocabulary. Ockey demonstrates how multiple-choice, true/false items, and short answer questions can be useful, though each has its limitations. Chapters 7–9 introduce item-writing guidelines and test design considerations for constructing fair and effective assessments. These chapters mainly focus on quantitative assessment methods such as descriptive statistics and item and content analyses that can be used to evaluate a test after it is administered. The author provides numerous examples of statistical techniques for gauging the quality and effectiveness of a test, such as mean, standard deviation, skewness, kurtosis, correlations, and Cronbach's Alpha.

The fourth part titled "Judging Test Taker Performances" includes three chapters on performance assessments requiring test takers to use the target language in communicative situations. This section also discusses validity evidence and approaches for evaluating the effectiveness of scores these performance assessments yield. Chapter 10 focuses on developing performance assessments, particularly speaking and writing tasks, and offers detailed examples. Chapter 11 demonstrates how performance assessments are rated using holistic and analytical rating scales and then addresses how human raters and computer systems can be trained to score performance assessments. In Chapter 12, the author examines types of validity evidence and sets forth guidelines for evaluating performance assessment tasks and estimating the dependability and reliability of scores. Intra-rater reliability is distinguished from inter-rater reliability, showing how consistency and fairness in test scores add pedagogical value.

The last part titled "Reflecting and Self-Assessment" consists of one concluding chapter encouraging readers to evaluate their own language assessment literacy and consider the ethical dimensions of their assessment practices. It integrates self-assessment tools and reflection questions that make it particularly suitable for use in language teacher education programs. Ockey emphasizes that language assessment is embedded in values, cultures, and institutional agendas. He urges L2 teachers to think critically about how to design and implement assessments that are more equitable, responsive, and impactful for learners' diverse needs.

Overall, this book stands out in several ways. Throughout the text, Ockey breaks down complex theoretical concepts into simple and clear explanations so that readers can easily comprehend the content and make effective use of L2 assessments. The use of real-life observations, experiences, and personal stories at the beginning of every chapter establishes a strong link between theory and practice and helps make the book more relatable than others in the language assessment literature.

To promote active learning, each chapter contains features such as reflective questions, hands-on activities, datasets for analysis, and discussion tasks. These elements can facilitate readers' understanding of how to design, implement, and evaluate assessments in the language classroom. Language teachers may appreciate the supplementary resources available with the book, which include PowerPoint slides, videos, datasets, and a comprehensive teacher's guide. These resources make it practical for teacher educators to integrate the book into their curricula.

Another noteworthy feature of the book is its integration of modern technological tools. Ockey addresses how computer-based testing, automated scoring systems, and generative AI are reshaping the landscape of language assessment. He specifies how the integration of computer scoring and hybrid scoring into assessment systems can minimize bias and increase the likelihood of fair evaluation of the test-taker, which may be particularly valuable for language teachers who struggle with developing fair and practical scoring tools. The author also acknowledges both the potential and pitfalls of using AI to create L2 test tasks. For instance, he discusses how AI might be used to develop test items, support scoring, and provide personalized feedback. At the same time, he raises concerns about transparency, fairness, and the need for human involvement—issues often overlooked in discussions of educational technology.

However, the book can be mildly critiqued for lacking detailed explanations about the use of AI. That is, generative AI is mentioned as an emerging tool in assessment, but the dearth of specific examples makes the discussion seem incomplete. Ockey only mentions the use of ChatGPT in constructing and rating assessments; a brief discussion of how other AI tools could be utilized in language assessment would have been helpful. Additionally, while Ockey provides instructions on how ChatGPT can be used for creating reading and writing tasks, he does not address how AI can be used to create rating scales.

Similarly, some readers may find that the discussion of statistical methods falls short. The statistical analyses might be difficult to understand for readers without basic knowledge of statistics. Although the author demonstrates how different statistical approaches are used in scoring systems and provides samples, some introductory explanation of the statistical concepts could have strengthened the text.

Nevertheless, these minor critiques can be overlooked. *Introducing Second Language Assessment* can serve not only as an introduction to the field of L2 assessment but also as a practical guide for creating, administering, scoring and judging tests. This book is a highly recommended resource for understanding the foundational and evolving principles of language assessment. Its integration of theoretical knowledge, practical insights, and contemporary concerns such as technology and ethics makes it valuable in today's educational landscape. Whether used as a textbook in a university course or as a self-study guide by language teachers, this book can provide the essential information about language assessment.

ZARIN TASNIM
Temple University